BOUNDS FOR CHARACTERISTIC ROOTS

by

MARVIN CARLTON PAPENFUSS

B. S., Winona State College, 1966

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Mathematics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1968

Approved by:

Leonard E. Fuller

Major Professor

TABLE OF CONTENTS

I. INTRODUCTION

To a pure mathematician, the zeros of the characteristic polynomial for a matrix are known as characteristic roots. To applied mathematicians, engineers, physicists, etc., they are known as eigenvalues, secular values, latent roots, or proper values. Whatever field of natural sciences one may be in, the characteristic roots of matrices play an important role.

Given a square matrix $A = (a_{k\lambda})$ of order n, where $a_{k\lambda}$ are elements of a field, any characteristic root w of A is a solution to the characteristic equation

$$[I,1] \qquad \det (A - wI) = 0.$$

Associated with any nonzero characteristic root w is a nonzero characteristic vector $X = (x_1, x_2, \ldots, x_n)'$ which is a non-trivial solution to the system of linear equations

$$AX = wX,$$

or, equivalently, the system

$$[I,2] \qquad \sum_{\lambda=1}^{n} a_{k\lambda}x_\lambda = wx_k, \; k = 1, 2, \ldots, n.$$

Due to the importance of this system of linear equations, [I,2], it will be referenced as such whenever it is used throughout this paper. In using this system of equations, we will sometimes be interested only in the two or occasionally three equations which involve the largest values for the x_i, say $|x_m|$ and $|x_p|$. However, when this is being done, it will be stated at that time.

It is the objective of this paper to organize some of the existing bounds for the characteristic roots of several types of matrices. In order to keep this paper to a reasonable length, and yet retain the continuity of the presentation, it will be necessary to state without proof some of the first-established bounds and also a few basic theorems on inequalities.

The two types of matrices to be considered are, first of all, matrices with arbitrary real or complex elements, and secondly, stochastic and generalized stochastic matrices.

Some of the bounds will be easily interpreted geometrically while others are either primarily theoretical or else give rise to another bound which is sharper and easier to apply.

Some bounds for the characteristic roots of an arbitrary square matrix of order n have been known for a long time. The first results that specifically gave bounds for the characteristic roots of a real matrix were due to Ivar Bendixson and are dated 1900. Then, shortly after 1900, A. Hirsch proved a theorem which established the first bound obtained for the characteristic roots of an arbitrary matrix with real or complex elements. Some of these first-established bounds will be given in this paper as a foundation for the development of some of the more recently obtained bounds.

Throughout this paper, the terms circle, disc and oval are used to mean both the closed curve and its interior. Thus, inequalities are used rather than equalities. However, the phrase "in or on" is used to prevent misunderstanding.

II. BACKGROUND AND BASIC BOUNDS

Let $A = (a_{k\lambda})$ be an arbitrary square matrix of order n. A. Hirsch proved the following theorem [11,1.3.1]:

Theorem (II.1). Any characteristic root w of A satisfies the inequality

$$|w| \leq n \max_{k,\lambda} |a_{k\lambda}|.$$

Later, I. Schur proved the following theorem [11,1.4.1]:

Theorem (II.2). If w_v denotes the v^{th} characteristic root of an arbitrary square matrix $A = (a_{k\lambda})$ of order n, then

$$\sum_{v=1}^{n} |w_v|^2 \leq \sum_{k,\lambda=1}^{n} |a_{k\lambda}|^2 .$$

Because Theorem (II.2) was arrived at after Theorem (II.1), it should contain a more precise bound. Indeed it does, since in Schur's theorem, one considers all $|a_{k\lambda}|$, not just max $|a_{k\lambda}|$.

For any arbitrary square matrix of order n with real or complex coefficients, we define

$$\sum_{\lambda=1}^{n} |a_{k\lambda}| = S_k, \ k = 1, 2,\ldots,n, \qquad \sum_{k=1}^{n} |a_{k\lambda}| = T_\lambda, \ \lambda = 1,2,\ldots,n,$$

and call these the k^{th} row and λ^{th} column sum respectively.

Theorem (II.3). For any non-zero characteristic root w of A,

$$|w|^2 \leq \max_{(k)} (S_k T_k).$$

For the proof, we shall consider two cases. For the first case, we assume that $S_k \neq 0$ for all k. After proving the theorem for $S_k \neq 0$, we then assume that $S_k = 0$ for some k, which actually means that the elements of the k^{th} row are zeros. We shall then consider the characteristic roots of a matrix similar to A in which the n^{th} row is the row of zeros. This will then imply that the n^{th} component of the characteristic vector associated with the root must be zero.

Proof. Assume $S_k \neq 0$ for all k.
Recall (I.2) that the basic system of linear equations for any characteristic root w is

$$wx_k = \sum_{\lambda=1}^{n} a_{k\lambda}x_\lambda .$$

By taking absolute values of both sides of this equality and applying the triangular inequality to the right side, we obtain

$$[II,1] \qquad |w||x_k| \leq \sum_{\lambda=1}^{n} |a_{k\lambda}||x_\lambda| = \sum_{\lambda=1}^{n} |a_{k\lambda}|^{1/2}|a_{k\lambda}|^{1/2}|x_\lambda| .$$

If we square both sides of [II,1], the inequality becomes

$$[II,2] \qquad |w|^2|x_k|^2 \leq \left(\sum_{\lambda=1}^{n} |a_{k\lambda}|^{1/2}|a_{k\lambda}|^{1/2}|x_\lambda|\right)^2 .$$

Applying the Cauchy-Schwartz inequality in the form of

$$\left(\sum_{\lambda=1}^{n} a_\lambda b_\lambda\right)^2 \leq \left(\sum_{\lambda=1}^{n} a_\lambda{}^2 \sum_{\lambda=1}^{n} b_\lambda{}^2\right)$$

to the right side of [II,2], we obtain

$$|w|^2|x_k|^2 \le \left(\sum_{\lambda=1}^{n} |a_{k\lambda}|\right)\left(\sum_{\lambda=1}^{n} |a_{k\lambda}||x_\lambda|^2\right),$$

and hence,

$$|w|^2|x_k|^2 \le S_k \sum_{\lambda=1}^{n} |a_{k\lambda}||x_\lambda|^2 . \tag{II,3}$$

Since $S_k \neq 0$ for all k, we can then divide the k^{th} inequality by S_k to obtain

$$|w|^2 \frac{|x_k|^2}{S_k} \le \sum_{\lambda=1}^{n} |a_{k\lambda}||x_\lambda|^2 .$$

Upon summing these over k, we will then introduce T_λ, obtaining

$$|w|^2 \sum_{k=1}^{n} \frac{|x_k|^2}{S_k} \le \sum_{\lambda=1}^{n} \sum_{k=1}^{n} |a_{k\lambda}||x_\lambda|^2 = \sum_{\lambda=1}^{n} T_\lambda |x_\lambda|^2.$$

Replacing the summation on λ by a summation on k, we have

$$|w|^2 \sum_{k=1}^{n} \frac{|x_k|^2}{S_k} \le \sum_{k=1}^{n} T_k |x_k|^2.$$

Since our summations are both over k, we can subtract the right side of the inequality from the left side, and combine the summations into one which leaves

$$\sum_{k=1}^{n} \left(\frac{|w|^2}{S_k} - T_k\right)|x_k|^2 \le 0.$$

Since $|x_k|^2 \ge 0$, there must be at least one value of k, say d, such that

$$\frac{|w|^2}{S_d} - T_d \leq 0$$

and thus

$$|w|^2 \leq S_d T_d \leq \max_{(k)} (S_k T_k) \quad \text{if } S_k \neq 0, \ k = 1, 2, \ldots, n.$$

Assume that $S_k = 0$ for one value of k, say k = i. Consider the matrix B obtained from A by a permutation of the i^{th} and n^{th} rows and columns. Since this permutation is a similarity transformation, B will have the same characteristic roots as A. With the assumption that $w \neq 0$, we must have the n^{th} component of the characteristic vector corresponding to w equal to zero. Thus, in inequality [II,3], we need to sum only to n-1, obtaining n-1 inequalities of the form:

$$|w|^2 |x_k|^2 \leq S_k \sum_{\lambda=1}^{n-1} |a_{k\lambda}| |x_\lambda|^2, \ k = 1, 2, \ldots, n-1.$$

Recall that $S_n = 0$ really means that $|a_{nk}| = 0$ for all k. Thus, for the matrix A, we may suppose $S_k \neq 0$, k = 1, 2, ..., n-1. We use the same argument as above to obtain

$$|w|^2 \leq \max_{(k)} (S_k \sum_{\lambda=1}^{n-1} |a_{\lambda k}|) = \max_{(k)} (S_k T_k).$$

If $S_i = 0$ and $S_j = 0$, we again consider the similarity transformations such that $S_n = 0$ and $S_{n-1} = 0$ and apply the same reasoning as above. Hence, by continuing this process, the theorem is proved.

Prior to the establishment of the bound in this theorem,

several other bounds were known which turn out to be special cases of this theorem. W. V. Parker stated and proved the following result [11; p. 144].

Theorem (II.4). If we let $S = \max_{(k)} \frac{(S_k + T_k)}{2}$, then

$$|w| \leq S.$$

This follows directly by considering

$$|w| \leq \max_{(k)} (S_k T_k)^{1/2} \leq \max_{(k)} \frac{(S_k + T_k)}{2} = S.$$

The first inequality is a direct consequence of Theorem (II.3), while the second is true because the geometric mean is less than or equal to the arithmetic mean for non-negative numbers.

A. B. Farnell stated and proved the following important theorem [11; p. 144].

Theorem (II.5). If $S = \max_{(k)} S_k$ and $T = \max_{(k)} T_k$, then

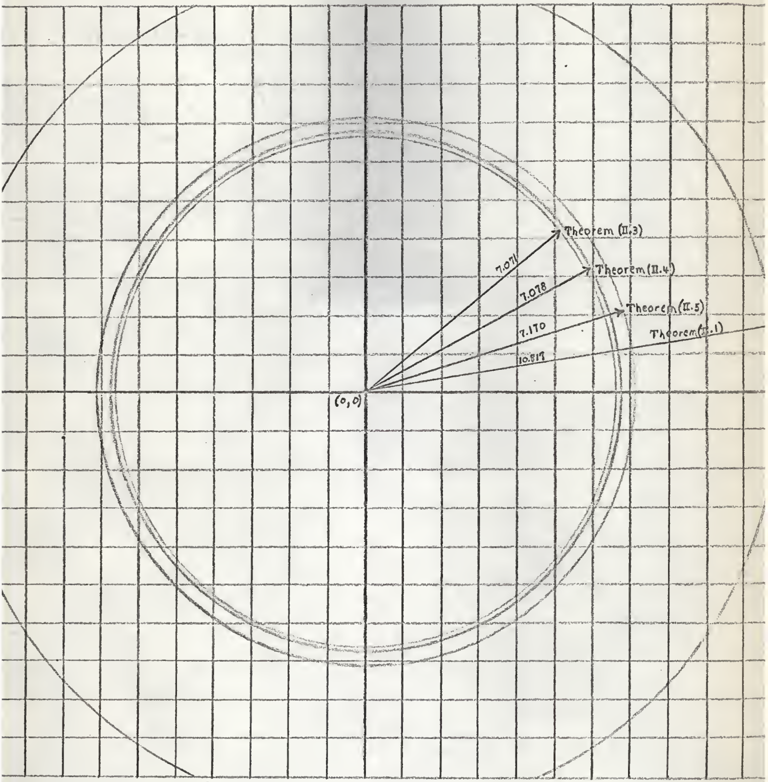
$$|w|^2 \leq ST.$$

This is a consequence of Theorem (II.3) since the maximum of a product is less than or equal to the product of the maxima. Although this bound is somewhat weaker than the bound from Theorem (II.3), it is, however, easier to apply.

Several of the circular bounds given thus far are illustrated in graph #1 for the matrix

$$A = \begin{bmatrix} 1+i & 2-3i & 1-2i \\ 3 & -2+i & 1 \\ 2-i & -1 & 1 \end{bmatrix}.$$

GRAPH #1


Theorem (II.3)
Theorem (II.4)
Theorem (II.5)
Theorem (II.1)
7.071
7.078
7.170
10.817
(0,0)

III. BOUNDS FOR THE CHARACTERISTIC ROOTS OF AN ARBITRARY SQUARE MATRIX

Let $A = (a_{k\lambda})$ be a square matrix with real or complex elements. Introduce

$$\sum_{\substack{\lambda=1 \\ \lambda \neq k}}^{n} |a_{k\lambda}| = P_k, \ k = 1, 2, \ldots, n.$$

Call this the k^{th} row radius.

Theorem (III.1). Each characteristic root w of A lies in the interior or on the boundary of at least one of the n circles

$$|z - a_{kk}| \leq P_k, \ k = 1, 2, \ldots, n.$$

Proof. For every non-zero characteristic root w, the system of linear equations, [I,2],

$$\sum_{\lambda=1}^{n} a_{k\lambda}x_\lambda = wx_k, \ k = 1, 2, \ldots, n,$$

has a non-trivial solution $(x_1, x_2, \ldots, x_n)$. Assume that $|x_v| \geq |x_j|$, $j = 1, 2, \ldots, n$, $j \neq v$. Then, consider the v^{th} equation of this system:

$$\sum_{\lambda=1}^{n} a_{v\lambda}x_\lambda = wx_v,$$

or equivalently, by transposing $a_{vv}x_v$ to the right side of the equation,

$$\sum_{\substack{\lambda=1 \\ \lambda \neq v}}^{n} a_{v\lambda}x_\lambda = (w - a_{vv})x_v \ .$$

Taking absolute values of both sides of the equation, and applying the triangle inequality, we have

$$|w - a_{vv}||x_v| = |\sum_{\substack{\lambda=1\\ \lambda\neq v}}^{n} a_{v\lambda}x_\lambda| \leq \sum_{\substack{\lambda=1\\ \lambda\neq v}}^{n} |a_{v\lambda}||x_\lambda| \leq \sum_{\substack{\lambda=1\\ \lambda\neq v}}^{n} |a_{v\lambda}||x_v|.$$

The last inequality follows since $|x_v| \geq |x_j|$ for all $j \neq v$. Since $(x_1, x_2, \ldots, x_n)$ was a non-trivial solution, $|x_v| \neq 0$. Thus

$$|w - a_{vv}| \leq \sum_{\substack{\lambda=1\\ \lambda\neq v}}^{n} |a_{v\lambda}| = P_v.$$

Hence w lies in the interior or on the boundary of at least one of the n circles $|z - a_{kk}| \leq P_k$, $k = 1, 2, \ldots, n$, and the theorem is proved. These closed circles are called Gersgorin discs.

Analogously, if we define

$$\sum_{\substack{k=1\\ k\neq\lambda}}^{n} |a_{k\lambda}| = Q_\lambda, \ \lambda = 1, 2, \ldots, n.$$

and call this the λ^{th} column radius, then each characteristic root w of A lies in the interior or on the boundary of at least one of the n circles

$$|z - a_{\lambda\lambda}| \leq Q_\lambda, \ \lambda = 1, 2, \ldots, n.$$

We now use this result to prove the following:

Theorem (III.2). The absolute value of each characteristic root is less than or equal to min(S, T), remembering that

$S = \max_{(k)} S_k$ and $T = \max_{(k)} T_k$.

Proof. From Theorem (III.1), we have the Gersgorin discs

$$|z - a_{kk}| \leq P_k \text{ and } |z - a_{\lambda\lambda}| \leq Q_\lambda.$$

Thus, for any characteristic root w,

$$|w - a_{kk}| \leq P_k \text{ and } |w - a_{\lambda\lambda}| \leq Q_\lambda.$$

If we apply the basic inequality that

$$|w - a_{kk}| \geq |w| - |a_{kk}|,$$

we obtain

$$|w| - |a_{kk}| \leq P_k \text{ and } |w| - |a_{\lambda\lambda}| \leq Q_\lambda.$$

By transposing, we complete the proof with

$$|w| \leq P_k + |a_{kk}| = S_k \text{ and } |w| \leq Q_\lambda + |a_{\lambda\lambda}| = T_\lambda.$$

Since k and λ were arbitrary, we obtain the result

$$|w| \leq \min(S, T).$$

Since $|a_{kk}| + P_k = S_k$, we have that all of the Gersgorin discs for the row radii lie within the circle $|z| \leq S$. Hence, if $|a_{kk}|$ is the minimum of the diagonal elements, then the Gersgorin disc centered at this a_{kk} will be tangent to the circle $|z| \leq S$. If $|a_{kk}| = 0$, then this Gersgorin disc will be the circle $|z| \leq s$.

A better bound than either the Gersgorin discs or the disc from Theorem (III.2) is presented in the next theorem.

The Gersgorin discs for both the row radii and the column

radii for the matrix displayed in section II are illustrated in graph #2, page 18. The shaded portion indicates the area contained in the union of the six Gersgorin discs, but not their intersection. The dash-constructed circles are for the row radii and the solid-constructed circles are for the column radii. The circle with center at (1, 0) is for both row and column radii. Since all of the characteristic roots of A must lie both in the union of the three discs using row radii and also in the union of the three discs using column radii, they must all lie in the intersection of all six discs. Graphically, this is the area enclosed by the shaded portion.

The largest circle, centered at the origin, is the bound from Theorem (III.2). Note that it encloses the intersection of the six Gersgorin discs, but not the union of them.

Theorem (III.3). Each characteristic root w of $A = (a_{k\lambda})$ lies in or on at least one of the $\binom{n}{2} = \frac{n(n - 1)}{2}$ ovals of Cassini:

$$|z - a_{kk}||z - a_{\lambda\lambda}| \le P_k P_\lambda, \; k,\lambda = 1, 2, \ldots, n, \; k \neq \lambda.$$

These ovals are a specialization of the generalized lemniscates. For a detailed geometric interpretation of these ovals, see [2, f].

Proof. As in Theorem (III.1), there will exist a non-zero characteristic vector $(x_1, x_2, \ldots, x_n)'$ corresponding to each non-zero characteristic root w which will be a non-trivial solution to the system of linear equations, [I,2],

$$\sum_{\lambda=1}^{n} a_{k\lambda}x_\lambda = wx_k, \ k = 1, 2, \ldots, n.$$

Let x_m and x_v be the two largest components in absolute value of $(x_1, x_2, \ldots, x_n)$, with

$$|x_m| \geq |x_v| \geq |x_i|, \ i = 1, 2, \ldots, n, \ i \neq m, v.$$

Consider now the m^{th} and the v^{th} equations of this system:

$$\sum_{\lambda=1}^{n} a_{m\lambda}x_\lambda = wx_m \text{ and } \sum_{\lambda=1}^{n} a_{v\lambda}x_\lambda = wx_v.$$

By transposing $a_{mm}x_m$ and $a_{vv}x_v$ respectively, these equations can equivalently be written as

$$\sum_{\substack{\lambda=1 \\ \lambda\neq m}}^{n} a_{m\lambda}x_\lambda = (w - a_{mm})x_m \text{ and } \sum_{\substack{\lambda=1 \\ \lambda\neq v}}^{n} a_{v\lambda}x_\lambda = (w - a_{vv})x_v.$$

Suppose $x_v = 0$. Then $x_i = 0$ for all $i \neq m$. Since $w \neq 0$, all of the x_i in the characteristic vector $(x_1, x_2, \ldots, x_n)'$ corresponding to w cannot be zero. Hence, $x_m \neq 0$. Thus, the m^{th} equation becomes

$$0 = (w - a_{mm})x_m = wx_m - a_{mm}x_m.$$

By transposing $a_{mm}x_m$ and then dividing both sides by the non-zero x_m, we get $w = a_{mm}$. Thus, w is trivially in the oval

$$|z - a_{mm}||z - a_{vv}| \leq P_mP_v.$$

If $x_v \neq 0$, we multiply both sides of the m^{th} equation by the

corresponding sides of the v^{th} equation, to obtain

$$(w - a_{mm})(w - a_{vv})x_m x_v = \sum_{\substack{\lambda=1 \\ \lambda\neq m}}^{n} a_{m\lambda}x_\lambda \sum_{\substack{\lambda=1 \\ \lambda\neq v}}^{n} a_{v\lambda}x_\lambda .$$

Upon taking absolute values of both sides and factoring out $|x_m|$ and $|x_v|$, the largest components of the $|x_i|$, from each summation, we have

$$|w - a_{mm}||w - a_{vv}||x_m x_v| \leq |x_v| \sum_{\substack{\lambda=1 \\ \lambda\neq m}}^{n} a_{m\lambda} \; |x_m| \sum_{\substack{\lambda=1 \\ \lambda\neq v}}^{n} a_{v\lambda} .$$

Now, applying the triangle inequality, we then introduce P_k as desired to obtain

[III,1] $$|w - a_{mm}||w - a_{vv}||x_m x_v| \leq |x_m| \sum_{\substack{\lambda=1 \\ \lambda\neq m}}^{n} |a_{m\lambda}| \; |x_v| \sum_{\substack{\lambda=1 \\ \lambda\neq v}}^{n} |a_{v\lambda}|$$

$$= |x_m|P_m|x_v|P_v .$$

The last equality is by the definition of the P_k, and hence, by dividing through by $|x_m||x_v|$, the inequality becomes

[III,2] $$|w - a_{mm}||w - a_{vv}| \leq P_m P_v .$$

Since $|x_m| \neq 0$ and $|x_v| \neq 0$, this division is permissible. Thus w lies in or on the oval

$$|z - a_{mm}||z - a_{vv}| \leq P_m P_v$$

and the theorem is proved.

If the column radii are used instead of the row radii, we get

a similar result. In this case, the statement is that w lies in or on at least one of the $\frac{n(n-1)}{2}$ ovals of Cassini

$$|z - a_{kk}||z - a_{\lambda\lambda}| \leq Q_k Q_\lambda, \; k,\lambda = 1, 2, \ldots, n, \; k \neq \lambda.$$

Suppose $a_{\lambda\lambda} < a_{kk}$. Then, it is true that any point which lies in the oval

$$|z - a_{kk}||z - a_{\lambda\lambda}| \leq Q_k Q_\lambda$$

also lies in the oval

$$|z - a_{\lambda\lambda}||z - a_{\lambda\lambda}| \leq Q_\lambda Q_\lambda.$$

But this latter oval is merely the Gersgorin disc

$$|z - a_{\lambda\lambda}| \leq Q_\lambda.$$

Hence, the above oval lies within the union of the two Gersgorin discs

$$|z - a_{kk}| \leq P_k \quad \text{and} \quad |z - a_{\lambda\lambda}| \leq P_\lambda,$$

and this is indeed a better result. If $k = \lambda$, these ovals become the Gersgorin discs.

These ovals of Cassini are somewhat difficult to construct because, to find any point on the curve, one must solve a fourth degree polynomial equation in x and y. However, these ovals are symmetric about the line joining their "foci", that is, the line joining the two diagonal entries being considered. They are also symmetric about the perpendicular bisector of the segment joining their "foci". Hence, four points on the oval are quite easily

found.

For the matrix displayed in section II, the ovals using the row radii are illustrated in graph #3. For the ovals

$$|z - (1 + i)||z - (-2 + i)| \leq (\sqrt{13} + \sqrt{5})4$$

and

$$|z - (1 + i)||z - 1| \leq (\sqrt{13} + \sqrt{5} + 1),$$

the four points on each curve are easy to find because their lines of symmetry are parallel to the x-axis and the y-axis. The equations for these lines, first for the lines joining their "foci", and then for the perpendicular bisectors of the segments joining their "foci" are, for the respective ovals,

$$y = 1 \quad \text{and} \quad x = -1/2$$

and

$$x = 1 \quad \text{and} \quad y = 1/2 \ .$$

However, for the oval

$$|z - (-2 + i)||z - 1| \leq 4(\sqrt{5} + 1),$$

the line joining their "foci" has the equation

$$x + 3y = 1,$$

and the perpendicular bisector of this segment joining (-2, 1) and (1, 0) has the equation

$$-3x + y = 2.$$

Hence, to find the four points on the curve, we must solve the

two systems of equations

$$(x^2 + 4x + y^2 - 2y + 5)^{1/2}(x^2 - 2x + y^2 + 1)^{1/2} = 4(\sqrt{5} + 1)$$

$$x + 3y = 1$$

and

$$(x^2 + 4x + y^2 - 2y + 5)^{1/2}(x^2 - 2x + y^2 + 1)^{1/2} = 4(\sqrt{5} + 1)$$

$$-3x + y = 2.$$

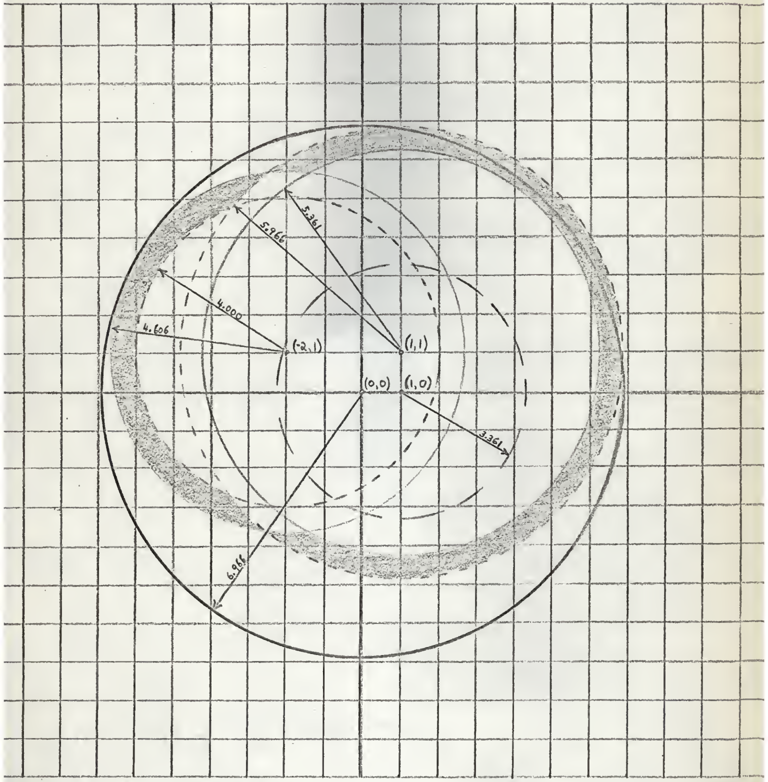
The solutions to these two pairs of equations were obtained on the IBM 360 computer and they are given on the graph to four decimal places.

As in the graph of the Gersgorin discs, the intersection of these three ovals of Cassini lies within the shaded portion. The reader is reminded that these are the three ovals obtained from the rows, and that the three other ovals obtained from the columns may even further decrease the area in which the characteristic roots for this particular matrix may lie. Since both graph #2 and graph #3 are to the same scale, the reader may hold the two graphs to the light and see that these ovals do indeed lie interior to the intersection of the Gersgorin discs.

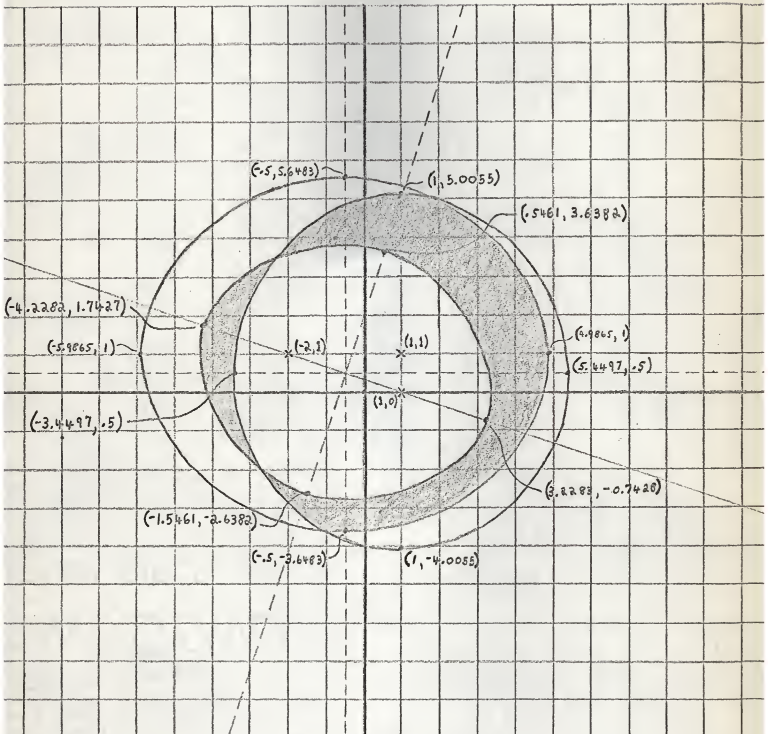
We shall now show that these bounds given by the ovals of Cassini can be improved. These improvements will be derived mathematically, and then shown, indeed, to be improvements. The first of these bounds, which is also a set of ovals, is contained in the next theorem.

Theorem (III.4). Let $A = (a_{k\lambda})$ be a square matrix of order n, and define

GRAPH #2



GRAPH #3


(-.5, 5.6483)
(1, 5.0055)
(.5461, 3.6382)
(-4.2282, 1.7427)
(-5.9865, 1)
(-2, 1)
(1, 1)
(4.9865, 1)
(5.4497, .5)
(1, 0)
(-3.4497, .5)
(3.2283, -0.7428)
(-1.5461, -2.6382)
(-.5, -3.6483)
(1, -4.0055)

$$P_k = |a_{k\lambda}|P_\lambda + |a_{\lambda k}|(P_k - |a_{k\lambda}|) + \sum_{v=1}^{n} |a_{kv}a_{\lambda v}| +$$

$$\sum_{v<\mu} |a_{kv}a_{\lambda\mu} + a_{k\mu}a_{\lambda v}|, \text{ where } \mu \neq k,\lambda,\ v \neq k,\lambda,\ k \neq \lambda.$$

Then each characteristic root of A lies in or on at least one of the ovals:

$$|z - a_{kk}||z - a_{\lambda\lambda}| \leq P_{k\lambda}$$

Proof. Let $(x_1, x_2, \ldots, x_n)$ be the non-trivial solution of the system of linear equations, [I,2], associated with w. Let

$$|x_m| \geq |x_v| \geq |x_i|,\ i = 1, 2, \ldots, n,\ i \neq m,\ i \neq v.$$

As in the proof of Theorem (III.3), consider the m^{th} and v^{th} equations in the form:

$$wx_m - a_{mm}x_m = \sum_{\substack{\lambda=1 \\ \lambda\neq m}}^{n} a_{m\lambda}x_\lambda$$

and

$$wx_v - a_{vv}x_v = \sum_{\substack{\lambda=1 \\ \lambda\neq v}}^{n} a_{v\lambda}x_\lambda.$$

If $x_v = 0$, we have $w = a_{mm}$ as in Theorem (III.3), so that w is at one of the "foci" of all ovals formed using a_{mm}. Hence, the inequality is trivially satisfied.

If $x_v \neq 0$, we have, after multiplying the respective sides of the two equations together,

$$(w - a_{mm})(w - a_{vv})x_m x_v = \sum_{\substack{\lambda=1 \\ \lambda\neq m}}^{n} a_{m\lambda}x_\lambda \sum_{\substack{\mu=1 \\ \mu\neq v}}^{n} a_{v\mu}x_\mu ,$$

which we can write as

$$[III,3] \qquad (w - a_{mm})(w - a_{vv})x_m x_v = a_{mv}x_v \sum_{\substack{\mu=1 \\ \mu\neq v}}^{n} a_{v\mu}x_\mu +$$

$$a_{vm}x_m \sum_{\substack{\lambda=1 \\ \lambda\neq v,m}}^{n} a_{m\lambda}x_\lambda + \sum_{\substack{\lambda=1 \\ \lambda\neq m,v}}^{n} a_{m\lambda}a_{v\lambda}x_\lambda^2 + \sum_{\substack{\mu\neq m,v \\ \lambda\neq m,v \\ \lambda<\mu}} (a_{m\lambda}a_{v\mu} + a_{m\mu}a_{v\lambda})x_\lambda x_\mu .$$

The various terms of the right side of equality [III,3] are obtained as follows:

Term 1. Remove $a_{mv}x_v$ from $\sum_{\substack{\lambda=1 \\ \lambda\neq m}}^{n} a_{m\lambda}x_\lambda$ and multiply it by $\sum_{\substack{\mu=1 \\ \mu\neq v}}^{n} a_{v\mu}x_\mu$.

Term 2. Remove $a_{vm}x_m$ from $\sum_{\substack{\mu=1 \\ \mu\neq v}}^{n} a_{v\mu}x_\mu$ and multiply it by $\sum_{\substack{\lambda=1 \\ \lambda\neq v,m}}^{n} a_{m\lambda}x_\lambda$.

Term 3. This is the sum of all products of two coefficients in the same position times the square of the x_λ in that position.

Term 4. This is the sum of two products, the first obtained by multiplying the element in the i^{th} position of $\sum_{\lambda=1}^{n} a_{m\lambda}x_\lambda$ by the

element in the j^{th} position of $\sum_{\mu=1}^{n} a_{v\mu} x_\mu$ and the second by multiplying the element in the j^{th} position of $\sum_{\lambda=1}^{n} a_{m\lambda} x_\lambda$ by the element in the i^{th} position of $\sum_{\mu=1}^{n} a_{v\mu} x_\mu$ and summing all of these.

By taking absolute values of both sides of [III,3], and applying the triangle inequality, we obtain the following inequality:

$$\text{[III,4]} \qquad |w - a_{mm}||w - a_{vv}||x_m x_v| \leq |a_{mv}||x_v| \sum_{\substack{\mu=1 \\ \mu\neq v}}^{n} |a_{v\mu}||x_\mu| + |a_{vm}||x_m| \sum_{\substack{\lambda=1 \\ \lambda\neq v,m}}^{n} |a_{m\lambda}||x_\lambda| + \sum_{\substack{\lambda=1 \\ \lambda\neq m,v}}^{n} |a_{m\lambda} a_{v\lambda}||x_\lambda|^2 + \sum_{\substack{\lambda<\mu \\ \lambda\neq m,v \\ \mu\neq m,v}} |a_{m\lambda} a_{v\mu} + a_{m\mu} a_{v\lambda}||x_\lambda x_\mu|.$$

For this proof, we assumed that

$$|x_m| \geq |x_v| \geq |x_i|, \; i = 1, 2, \ldots, n.$$

Thus, without loss of generality, we can normalize this vector on any $x_i \neq 0$, $i \neq m$, making this component unity. Hence, from terms 2 and 3, the following inequalities are true:

$$\sum_{\substack{\mu=1 \\ \mu\neq v}}^{n} |a_{v\mu}||x_\mu| \leq \sum_{\substack{\mu=1 \\ \mu\neq v}}^{n} |a_{v\mu}| \quad \text{and} \quad \sum_{\substack{\lambda=1 \\ \lambda\neq v,m}}^{n} |a_{m\lambda}||x_\lambda| \leq \sum_{\substack{\lambda=1 \\ \lambda\neq v,m}}^{n} |a_{m\lambda}|.$$

On using the notation previously described for row radii, [III,4] becomes

$$[\text{III},5] \qquad |w - a_{mm}||w - a_{vv}||x_m x_v| \leq |a_{mv}||x_v|P_v + |a_{vm}||x_m|(P_m - |a_{mv}|) + \sum_{\substack{\lambda=1 \\ \lambda\neq m,v}}^{n} |a_{m\lambda}a_{v\lambda}||x_\lambda|^2 + \sum_{\substack{\lambda<\mu \\ \lambda\neq m,v \\ \mu\neq m,v}} |a_{m\lambda}a_{v\mu} + a_{m\mu}a_{v\lambda}||x_\lambda x_\mu|.$$

The above normalizing also yields the following set of inequalities for λ, $\mu < m$:

$$[\text{III},6] \qquad |x_m x_v| \geq |x_v|, \quad |x_m x_v| \geq |x_m|,$$

$$|x_m x_v| \geq |x_\lambda|^2, \quad |x_m x_v| \geq |x_\lambda x_\mu|.$$

Thus, upon dividing both sides of inequality [III,5] by $|x_m x_v|$, the inequality remains valid and can be written as

$$[\text{III},7] \qquad |w - a_{mm}||w - a_{vv}| \leq |a_{mv}|P_v + |a_{vm}|(P_m - |a_{mv}|) + \sum_{\substack{\lambda=1 \\ \lambda\neq m,v}}^{n} |a_{m\lambda}a_{v\lambda}| + \sum_{\substack{\lambda<\mu \\ \lambda\neq m,v \\ \mu\neq m,v}} |a_{m\lambda}a_{v\mu} + a_{m\mu}a_{v\lambda}| = P_{mv}.$$

Hence w lies in or on at least one of the ovals

$$|z - a_{mm}||z - a_{vv}| \leq P_{mv},$$

which proves the theorem.

To verify that $P_{mv} \leq P_m P_v$, we note that in the proof of Theorem (III.3), involving $P_m P_v$, the last operation was to divide

both sides of the inequality [III,1] by $|x_m x_v|$. However, this quantity appeared on both sides of the inequality, so it divided to unity. On the other hand, in the proof of Theorem (III.4), involving P_{mv}, when we divided both sides of inequality [III,5] by $|x_m x_v|$, we obtained unity on the left, but not on the right unless, in fact, the inequalities in [III,6] were actually all equalities. Hence, the ovals of Theorem (III.4) are at least as good as, and will in general be better than, the original ovals of Cassini obtained in Theorem (III.3).

One might ask whether or not these ovals have any practical advantage over the Gersgorin discs, and here is one of the reasons why. First, note that since the Gersgorin discs are easily constructed, these should be the first bounds to consider when investigating the location of the characteristic roots of a given matrix. However, if the bounds obtained by using them are still not sufficiently accurate, we then resort to using the ovals of Cassini. As has been stated, the ovals lie within the union of the Gersgorin discs, and thus, for instance, if one questions whether or not a point near the boundary of a disc is an upper bound for a characteristic root, it may be confirmed by using the ovals.

Then too, it will not always be necessary to consider all of the $\frac{n(n-1)}{2}$ ovals. For example, if we know that a root lies within the circle $|z - a_{11}| \leq P_1$, then we need only consider the $(n - 1)$ ovals

$$|z - a_{11}||z - a_{\lambda\lambda}| \leq P_1 P_\lambda, \ \lambda = 2, 3, \ldots, n.$$

Furthermore all of these may not need to be considered if sufficient bounds are obtained prior to the $(n - 1)^{st}$ oval.

If we specialize our matrix and require that $A = (a_{k\lambda})$ be a square matrix of order n with real elements, we can obtain the following theorem, which is very similar to Theorem (III.4). This theorem and its formal proof can be found in [2, e, p.557].

Theorem (III.5). Let $A = (a_{k\lambda})$ be a square matrix of order n with real elements. For each given k and λ we denote the sum of the positive terms of

$$S_{k\lambda} = \sum_{\substack{v=1 \\ v\neq k,\lambda}}^{n} a_{kv}a_{\lambda v}$$

by $U_{k\lambda}$, and the sum of the negative terms of this $S_{k\lambda}$ by $V_{k\lambda}$, and denote $\max(U_{k\lambda}, |V_{k\lambda}|)$ by $m_{k\lambda}$. In a similar fashion as we did in Theorem (III.4), we set

$$P^*_{k\lambda} = |a_{k\lambda}|P_\lambda + |a_{\lambda k}|(P_k - |a_{k\lambda}|) + m_{k\lambda} + \sum_{v<\mu} |a_{kv}a_{\lambda\mu} + a_{k\mu}a_{\lambda v}|.$$

Then each real characteristic root w of A must lie in at least one of the closed intervals formed by the $\frac{n(n - 1)}{2}$ ovals

$$|z - a_{kk}||z - a_{\lambda\lambda}| \leq P^*_{k\lambda},$$

and the real axis.

The proof is very similar to the proof of Theorem (III.4). The only difference between the two theorems is that the third term of $P_{k\lambda}$, namely,

$$\sum_{v=1}^{n} |a_{kv}a_{\lambda v}|,$$

is replaced by $m_{k\lambda} = \max (U_{k\lambda}, |V_{k\lambda}|)$ in $P^*_{k\lambda}$.

We see that $m_{k\lambda}$ is either the sum of the positive terms of $S_{k\lambda}$ or the sum of the absolute values of the negative terms of $S_{k\lambda}$, whichever is larger. Thus,

$$m_{k\lambda} \leq \sum_{v=1}^{n} |a_{kv}a_{\lambda v}|,$$

and this gives us

$$P^*_{k\lambda} \leq P_{k\lambda}.$$

Hence, Theorem (III.5) gives a sharper bound than Theorem (III.4) did, but we must remember that this latter theorem can be applied only to the real characteristic roots of a matrix with real elements.

IV. BOUNDS FOR THE CHARACTERISTIC ROOTS OF STOCHASTIC AND GENERALIZED STOCHASTIC MATRICES

All of the bounds obtained so far with the exception of the bound obtained in Theorem (III.5) have been for arbitrary square matrices of order n with real or complex elements. Suppose now that we restrict the matrices under consideration in different ways. For the following development, all elements of the matrix $A = (a_{k\lambda})$ are assumed to be non-negative.

We call a square matrix $A = (a_{k\lambda})$ of order n stochastic if

[IV,1] $$\sum_{\lambda=1}^{n} a_{k\lambda} = 1, \; k = 1, 2, \ldots, n,$$

and positive stochastic if $a_{k\lambda} \neq 0$ for all k and λ. This definition is extended by calling a square matrix $A = (a_{k\lambda})$ of order n generalized stochastic if

[IV,2] $$\sum_{\lambda=1}^{n} a_{k\lambda} = g, \; k = 1, 2, \ldots, n,$$

where g is some constant, and positive generalized stochastic if $a_{k\lambda} \neq 0$ for all k and λ.

By looking at the defining system of linear equations for a characteristic root, [I,2],

$$\sum_{\lambda=1}^{n} a_{k\lambda} x_{\lambda} = w x_k,$$

we see that for a stochastic matrix, $w = 1$ is a characteristic root and $(1, 1, \ldots, 1)'$ is a characteristic vector corresponding to $w = 1$.

For a stochastic matrix, all of the row sums are one. Hence, as pointed out in the comment following the proof of Theorem (III.2), all of the characteristic roots of a stochastic matrix must lie in or on the unit circle $|z| \leq 1$. Since $S = \max_k S_k = 1$, the Gersgorin disc will be the unit circle if and only if $a_{\lambda\lambda} = 0$ for a particular λ. Otherwise, since $|a_{kk}| + P_k = S_k = 1$, the Gersgorin discs will all lie within the unit circle, and they will be tangent to it at exactly one point. This is the point of intersection of the line joining a_{kk} and the origin with the unit circle. This is also a consequence of the next theorem to be proved, Theorem (IV.1), but the result is already intuitively obvious.

Another carry-over to stochastic matrices is the following.

Theorem (IV.1). If $a_{kk} = \min_{(i)} a_{ii}$, $i = 1, 2, \ldots, n$, for any stochastic matrix $A = (a_{k\lambda})$, then all the characteristic roots lie in or on the circle

[IV,3] $$|z - a_{kk}| \leq 1 - a_{kk}.$$

Proof. This reduces to the unit circle if $a_{kk} = 0$.

If no $a_{kk} = 0$, this becomes the circle with center at a_{kk} with the largest possible radius, $1 - a_{kk}$, of any of the Gersgorin discs, and will include all of the other discs with centers at $a_{\lambda\lambda}$, $\lambda = 1, 2, \ldots, n$, $\lambda \neq k$, and with smaller radii $1 - a_{\lambda\lambda}$. Hence, all of the roots will lie in or on the circle [IV,3].

Instead of having $\frac{n(n-1)}{2}$ ovals of Cassini to consider in obtaining bounds for the characteristic roots, we shall prove

that only one oval is needed which will contain all of the roots of a stochastic matrix. The following two lemmas, stated without proof, will be needed in the development. For their proofs, see [2, d, pp. 78-80].

Lemma (IV.1). Let a, b, c and k be real numbers satisfying $a < c < k$ and $b \leq c < k$. Then the oval

$$|z - b||z - c| \leq (k - b)(k - c)$$

lies in the interior of the oval

$$|z - a||z - b| \leq (k - a)(k - b),$$

and $z = k$ is the only common point on the contours of both regions.

Lemma (IV.2). Assume that $a_1 \leq a_2 \leq \ldots \leq a_n < k$. Each of the ovals

$$|z - a_p||z - a_\lambda| \leq (k - a_p)(k - a_\lambda), \tag{IV,4}$$

$p, \lambda = 1, 2, \ldots, n$, $p < \lambda$, is either identical with the oval

$$|z - a_1||z - a_2| \leq (k - a_1)(k - a_2) \tag{IV,5}$$

or lies in the interior of the oval [IV,5]. The point $z = k$ is the only common point of the boundaries of the two different ovals [IV,4].

Assuming these lemmas, we are now in a position to apply them to a stochastic matrix in obtaining an oval which will contain all of the characteristic roots.

Theorem (IV.2). Let $A = (a_{k\lambda})$ be a stochastic matrix of order n. Let

$$a_{rr} \leq a_{tt} < a_{ii},\ i = 1, 2, \ldots, n,\ i \neq r, t.$$

Then, all of the characteristic roots of A lie in or on the single oval

$$[IV,6] \qquad |z - a_{rr}||z - a_{tt}| \leq (1 - a_{rr})(1 - a_{tt}).$$

Proof. The proof is a direct consequence of our previous Theorem (III.3) on the ovals of Cassini combined with Lemma (IV.2). From Theorem (III.3), we know that all of the characteristic roots of A lie in or on at least one of the $\frac{n(n-1)}{2}$ ovals of Cassini

$$|z - a_{kk}||z - a_{\lambda\lambda}| \leq P_k P_\lambda,\ k, \lambda = 1, 2, \ldots, n,\ k \neq \lambda.$$

But, for A a stochastic matrix,

$$P_k = (1 - a_{kk}) \text{ and } P_\lambda = (1 - a_{\lambda\lambda}).$$

Hence, all of the characteristic roots lie in or on at least one of the $\frac{n(n-1)}{2}$ ovals

$$[IV,7] \qquad |z - a_{kk}||z - a_{\lambda\lambda}| \leq (1 - a_{kk})(1 - a_{\lambda\lambda}),$$

for $k, \lambda = 1, 2, \ldots, n$, $k \neq \lambda$. If we apply Lemma (IV.2) to the elements of A, where k = 1, we have

$$a_1 \leq a_2 \leq \cdots \leq a_n < 1.$$

Applying Lemma (IV.2) to the diagonal elements of A in a very slightly modified form, we have

$$a_{rr} \leq a_{tt} < a_{ii} \leq 1.$$

Thus, since a_{rr} and a_{tt} are less than one, we have each of the ovals of [IV,7] lying within the single oval of [IV,6]. Hence, the theorem is proved.

The result of this theorem can easily be carried over to a generalized stochastic matrix.

Theorem (IV.3). If $A = (a_{k\lambda})$ is a generalized stochastic matrix with row sum g, and if

$$a_{rr} \leq a_{tt} < a_{ii},\ i = 1, 2, \ldots, n,\ i \neq r, t,$$

then all of the characteristic roots of A lie in or on the single oval

$$|z - a_{rr}||z - a_{tt}| \leq (g - a_{rr})(g - a_{tt}).$$

The proof of this is similar to the proof of Theorem (IV.2) and is therefore omitted.

Recall that for a stochastic matrix w = 1 is a trivial characteristic root. Similarly, for a generalized stochastic matrix with row sum g, w = g is a trivial characteristic root. Note that w = 1 will lie on the boundary of the oval

$$|z - a_{rr}||z - a_{tt}| \leq (1 - a_{rr})(1 - a_{tt}),$$

where

$$a_{rr} \leq a_{tt} \leq a_{ii},\ i = 1, 2, \ldots, n,\ i \neq r, t.$$

Also, for the generalized stochastic matrix, w = g will lie on

the boundary of the oval

$$|z - a_{rr}||z - a_{tt}| \le (g - a_{rr})(g - a_{tt}).$$

But, what can be said about the non-trivial characteristic roots of a stochastic or generalized stochastic matrix? Are there smaller bounds which will necessarily include all of these roots? These questions are answered with the following theorem, stated without the proof, which can be found in [2, d, p. 89].

Theorem (IV.4). Assume that m is the smallest off-diagonal element of the positive stochastic matrix $A = (a_{k\lambda})$ of order n and that a_{11} and a_{22} are the smallest elements of the main diagonal. Then all the non-trivial characteristic roots lie in or on the oval

$$|z - (a_{11} - m)||z - (a_{22} - m)| \le$$

$$\{1 - a_{11} - (n-1)m\}\{1 - a_{22} - (n-1)m\}.$$

Note that this theorem will also be true if A is a stochastic matrix rather than a positive stochastic matrix. For this case m would equal zero, and this would then be the same bound we had from Theorem (IV.2). But, $m \neq 0$ and hence the oval from Theorem (IV.4) is strictly smaller than the oval from Theorem (IV.2), and the trivial root $w = 1$ will indeed lie outside this new oval.

If we extend Theorem (IV.4) to generalized stochastic matrices, we obtain:

Theorem (IV.5). Assume that m is the smallest off-diagonal

element of the positive generalized stochastic matrix $A = (a_{k\lambda})$ with row sum g, and that a_{11} and a_{22} are the smallest elements of the main diagonal. Then all of the non-trivial characteristic roots lie in or on the oval

$$|z - (a_{11} - m)||z - (a_{22} - m)| \leq$$

$$\{g - a_{11} - (n-1)m\}\{g - a_{22} - (n-1)m\}.$$

Obviously, the above five theorems on stochastic matrices comprise merely an introduction to the theory of the localization of their characteristic roots. However, these bounds do serve as rudiments which the interested reader may incorporate into his further study on the subject. Such a study of localization theory for the characteristic roots of stochastic matrices may be done in many fields, and in particular, the field of probability and mathematical statistics, where it is extremely important in applications.

V. CONCLUSION

As has been stated earlier, one may easily construct some of the elementary bounds for the characteristic roots of a matrix. However, in order to improve upon these bounds, one must, in general, sacrifice both ease of computation and simplicity of construction. This continues until the geometric interpretation of a bound becomes almost impossible to visualize. For example, the bounds obtained from Theorem (III.4) and (III.5) are of this nature.

The order of presentation of the bounds in this paper has been determined by two factors. These are the increasing of the accuracy of the bound, and the decreasing of the ease of computation of the bound. It is true that these factors coincide for most of the bounds, as stated above. Nevertheless, there are instances in the paper where they differ. The bound from Theorem (II.5) is of this nature. From Graph #1 it is seen that the disc from Theorem (II.5) is larger than the disc from Theorem (II.3), but it is much easier to calculate the maximum S_k and the maximum T_k and then use this product as the radius of the disc than it is to calculate the maximum $S_k T_k$ for all k.

If one is posed with a practical problem, such as may arise in the construction of a bridge or other solid structure, where an upper bound on the characteristic roots of a particular matrix is desired, which bound should he use? The answer to this question depends upon many factors, and probably the most important factor is: will the structure be built to meet only the minimum

requirements for strength and durability or will it be built to withstand a stress much greater than the maximum stress it is expected to receive? If the former is the case, one would probably want a very accurate bound, and thus, the ovals of Cassini from Theorem (III.3) may suffice. If a still better bound is needed, Theorem (III.4) or Theorem (III.5) may be used. On the other hand, if the latter is the case, a somewhat rougher bound may be sufficient. In this case, a good one to use is the disc from Theorem (III.2) since it is easy to apply. Although this bound may be easier to apply, it may not be a sufficiently accurate bound. Hence, one may choose to use the Gersgorin discs from Theorem (III.1).

Therefore, depending upon the desired degree of accuracy needed for a given situation, one may choose the bound which is best suited.

ACKNOWLEDGMENT

The writer wishes to express his sincere appreciation to Dr. Leonard E. Fuller for his patient guidance and supervision given during the preparation of this report.

REFERENCES

1. Barankin, E. W., "Bounds for the Characteristic Roots of a Matrix", Bulletin of the American Mathematical Society, Vol. 51 (1945), pp. 767-770.

2. Brauer, Alfred, "Limits for the Characteristic Roots of a Matrix",
   a) Duke Mathematical Journal, Vol. 13 (1946), pp. 387-95;
   b) II, Vol. 14 (1947), pp. 20-26;
   c) III, Vol. 15 (1948), pp. 871-877;
   d) IV, Vol. 19 (1952), pp. 75-91;
   e) V, Vol. 19 (1952), pp. 553-562;
   f) VII, Vol. 25 (1958), pp. 583-590.

3. Brauer, Alfred, "The Theorems of Ledermann and Ostrowski on Positive Matrices", Duke Mathematical Journal, Vol. 24 (1957), pp. 265-274.

4. Brauer, Alfred, "A New Proof of Theorems of Perron and Frobenius on Non-negative Matrices: I. Positive Matrices", Duke Mathematical Journal, Vol. 24 (1957), pp. 367-378.

5. Brauer, Alfred, "Stochastic Matrices with a Non-trivial Greatest Positive Root", Duke Mathematical Journal, Vol. 27 (1960), pp. 291-296.

6. Brauer, Alfred, "On the Characteristic Roots of Power-positive Matrices", Duke Mathematical Journal, Vol. 28 (1961), pp. 439-446.

7. Fuller, Leonard E., Matrix Theory. Presently unpublished. pp. 7-1 to 7-25.

8. Gantmacher, F. R., Applications of the Theory of Matrices, Interscience Publishers, Inc., New York, London, pp. 75-79, pp. 187-194.

9. Householder, Alston S., The Theory of Matrices in Numerical Analysis, Blaisdell Publishing Company, New York, London, Toronto, pp. 37-85.

10. Ledermann, Walter, "Bounds for the Greatest Latent Root of a Matrix", Journal of the London Mathematical Society, Vol. 25 (1950), pp. 265-268.

11. Marcus, Marvin, and Minc, Henry, A Survey of Matrix Theory and Matrix Inequalities, Allyn and Bacon, Inc., Boston, 1964, pp. 139-169.

12. Ostrowski, Alexander, "Bounds for the Greatest Latent Root of a Positive Matrix", Journal of the London Mathematical Society, Vol. 27 (1952), pp. 253-256.

13. Ostrowski, Alexander, and Schneider, Hans, "Bounds for the Maximal Characteristic Root of a Non-negative Irreducible Matrix", Duke Mathematical Journal, Vol. 27 (1960), pp. 547-553.

14. Schneider, Hans, Recent Advances in Matrix Theory, The University of Wisconsin Press, Madison and Milwaukee, 1964, pp. 3-38.

15. Taussky, Olga, "Bounds for Characteristic Roots of Matrices", Duke Mathematical Journal, Vol. 15 (1948), pp. 1043-1044.

16. Varga, Richard S., Matrix Iterative Analysis, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1962, pp. 16-34.

BOUNDS FOR CHARACTERISTIC ROOTS

by

MARVIN CARLTON PAPENFUSS

B. S., Winona State College, 1966

AN ABSTRACT OF A REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Mathematics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1968

The theory on bounds for the characteristic roots of a matrix may be classified as a branch of the theory of inequalities. Both upper bounds and lower bounds can be found, but we shall restrict our consideration to upper bounds only. These bounds are derived as geometric configurations. Some are circles and some are ovals.

The k^{th} row sum of a square matrix of order n is defined as

$$S_k = \sum_{\lambda=1}^{n} |a_{k\lambda}|$$

where $k = 1, 2, \ldots, n$. Similarly, the λ^{th} column sum is defined as

$$T_\lambda = \sum_{k=1}^{n} |a_{k\lambda}| ,$$

where $\lambda = 1, 2, \ldots, n$.

The first important circular bound established is that all of the characteristic roots w of an arbitrary square matrix of order n lie in or on the circle

$$|z| \leq \{(\max_k S_k)(\max_\lambda T_\lambda)\}^{1/2} .$$

This bound is both proved mathematically and depicted graphically.

Using this same notation for row sums and column sums, it is then proved that all of the roots lie in or on the circle

$$|z| \leq \max_k \frac{S_k + T_k}{2} .$$

The k^{th} row radius, P_k, is defined as $P_k = S_k - |a_{kk}|$. Similarly, the λ^{th} column radius, Q_λ, is defined as $Q_\lambda = T_\lambda - |a_{\lambda\lambda}|$.

For a square matrix of order n, the n discs with centers at

a_{kk} and radii of P_k are called the Gersgorin discs for the rows of the matrix. If the radii are the Q_k, they are the Gersgorin discs for the columns of the matrix.

It is proved that all of the characteristic roots of a square matrix lie in or on both the union of the n Gersgorin discs for the rows and also the union of the n Gersgorin discs for the columns. Hence, they lie in the intersection of the 2n discs.

If the row radii are taken in pairs, the bound established is that all of the characteristic roots w lie in or on the $\frac{n(n-1)}{2}$ ovals of Cassini

$$|z - a_{kk}||z - a_{\lambda\lambda}| \leq P_k P_\lambda.$$

A similar set of $\frac{n(n-1)}{2}$ ovals of Cassini for the columns is

$$|z - a_{kk}||z - a_{\lambda\lambda}| \leq Q_k Q_\lambda.$$

Thus, all of the roots lie within the intersection of these $n(n - 1)$ ovals.

These ovals are proved to be better than the Gersgorin discs. The intersection of the $\frac{n(n-1)}{2}$ ovals for the rows are depicted graphically, and a considerable improvement over the Gersgorin discs can be observed.

By making use of several stated lemmas, it is proved that all of the characteristic roots of a stochastic matrix lie in or on the single oval

$$|z - a_{rr}||z - a_{tt}| \leq (1 - a_{rr})(1 - a_{tt}),$$

where a_{rr} and a_{tt} are the two smallest diagonal elements.

A smaller oval than the above is defined in terms of the two smallest diagonal elements, say a_{11} and a_{22}, and the minimum off-diagonal element, say m. The result is that all of the characteristic roots of a stochastic matrix of order n lie in or on the oval

$$|z - (a_{11} - m)||z - (a_{22} - m)| \leq \{1 - a_{11} - (n-1)m\}\{1 - a_{22} - (n-1)m\}.$$

All of the results for stochastic matrices are also proved for generalized stochastic matrices where 1 is replaced by 2, the constant row sum.